




Head-EyeK: Head-eye Coordination and Control Learned in Virtual Reality

Yifang Pan , Ludwig Sidenmark , and Karan Singh 

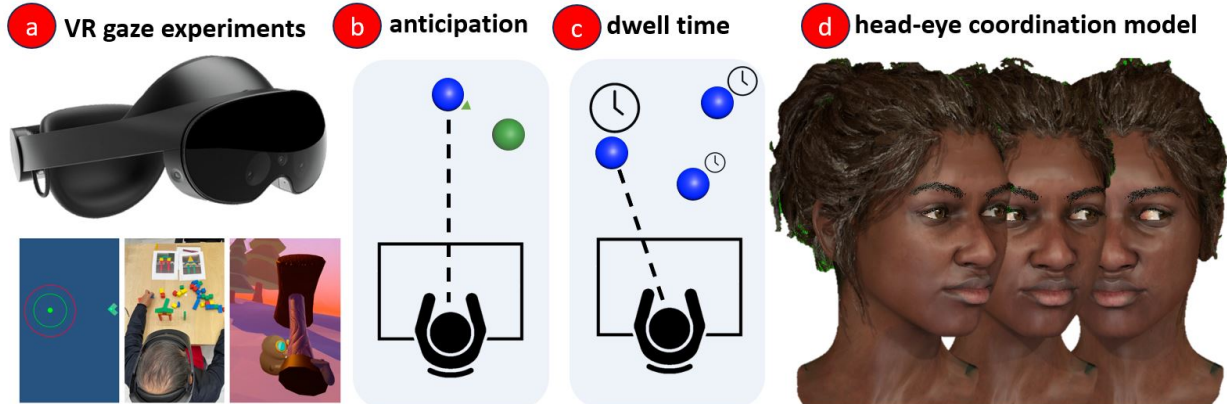


Fig. 1: We designed multiple experiments in Virtual and Augmented Reality to collect head-eye coordination data for sequences of gaze motion (a). We are the first to model head-eye coordination behavior that considers the effect of anticipated gaze (b), and intended dwell times (c) of each gaze fixation. We generate convincing coordinated head and eye movements that match collected ground truth data (d).

Abstract—Human head-eye coordination is a complex behavior, shaped by physiological constraints, psychological context, and gaze intent. Current context-specific gaze models in both psychology and graphics fail to produce plausible head-eye coordination for general patterns of human gaze behavior. In this paper, we: 1) propose and validate an experimental protocol to collect head-eye motion data during sequential look-at tasks in Virtual Reality; 2) identify factors influencing head-eye coordination using this data; and 3) introduce a head-eye coordinated Inverse Kinematic gaze model *Head-EyeK* that integrates these insights. Our evaluation of *Head-EyeK* is three-fold: we show the impact of algorithmic parameters on gaze behavior; we show a favorable comparison to prior art both quantitatively against ground-truth data, and qualitatively using a perceptual study; and we show multiple scenarios of complex gaze behavior credibly animated using *Head-EyeK*.

Index Terms—Virtual Humans, Computer Graphics Techniques.

1 INTRODUCTION

Our gaze serves as our primary means of interacting with the world and is a crucial element of communication, relying on the coordinated efforts of the eyes, head, and body. As such, being able to accurately model human gaze behavior is a key interest for virtual reality (VR) as it enables creating realistic and convincing virtual avatars [1], allowing accurate foveated rendering [2], improving ergonomics [3, 4], and predicting and enhancing interaction [5, 6]. Although the mechanics of eye and head movements have been extensively studied in behavioral psychology [7, 8], modeling the coordination between head and eye movements remains a less explored area of research. A key question arises for modeling the eye-head relationship: How much should the head rotate to accommodate the movement of the eyes?

Many existing models of gaze behavior fail to account for the continuous and dynamic nature of gaze shifts in real-world settings, where the gaze constantly shifts between multiple targets. These models often treat gaze shifts as isolated events originating from a neutral position

and do not consider that the body treats different gaze shifts differently [9–13]. Factors such as dwell time—the duration spent fixating on a target—and the presence of subsequent targets significantly influence how the head and eyes coordinate during gaze shifts [14]. As a result, when applied to real-world gaze data, these models tend to produce unnatural head movements, particularly when simulating sequences of gaze shifts, because they neglect past states (e.g., previous head and gaze positions) and future intentions (e.g., planned gaze shifts to upcoming targets).

To address these limitations and more accurately model head-eye coordination behaviors, we introduce *Head-EyeK*, a novel model that simulates natural head and eye movements during sequential gaze tasks. Building on Oommen’s [14] findings that the degree of head rotation towards a look-at point depends on both the gaze dwell time and the location of the next target, we designed a VR replica of their experimental setup. We expanded the original experiments to gain further insight into sequential gaze behaviors, which informed the development of *Head-EyeK*. Our model accounts for additional contexts such as past head and gaze positions and future gaze intentions, enabling more realistic simulations of continuous gaze shifts.

Our work validates the findings of [14], confirming that for look-at points with short dwell times, minimal head rotation occurs, while longer dwell times result in increased head rotation. We also demonstrate that the need to attend to a secondary target influences how much the head turns toward the primary target. These insights were incorporated into an algorithm for generating coordinated head-eye gaze animations, and both quantitative and qualitative evaluations show that

• Yifang Pan, Ludwig Sidenmark, and Karan Singh are with the University of Toronto.

• E-mail: evan.pan@mail.utoronto.ca, l.sidenmark@utoronto.ca, karan@dgp.toronto.edu

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

our approach produces significantly more natural animations compared to prior works.

Our contributions are as follows:

- Experimental evidence showing that dwell time significantly affects the extent of head movement during sequential gaze tasks, and anticipation of future gaze targets influences the head’s movement towards the current target.
- The *Head-EyeK* model, which anticipates head movement based on consecutive gaze targets and the time spent on each gaze target.
- Results from a comparative study of modeling gaze animations showing the increased accuracy of Head-EyeK over previous bio-mechanically-based head-eye coordination methods.

2 NOTATION AND BACKGROUND

Our vision is sharpest in our fovea, a 5° visual angle around the center of our field of view [15]. We thus constantly shift our gaze to bring different objects of interest into focus, known as foveation. We control this gaze using both our head and eyes. We express the overall gaze direction as $\mathbf{g} = \mathbf{h} + \mathbf{e}$, where \mathbf{g} , \mathbf{h} , and \mathbf{e} are 2D vectors representing yaw and pitch angles in degrees. While \mathbf{g} and \mathbf{h} represents gaze and head in world space, \mathbf{e} represents the eye in the local space of the head. We justify a 2D $\{h_{pitch}, h_{yaw}\}$ head model by noting that head roll only marginally affects gaze direction, and that the head physiologically follows Donder’s law [16], asserting a 1-1 relationship between a $(\{h_{pitch}, h_{yaw}, h_{roll}\})$ 3D head and our $\{h_{pitch}, h_{yaw}\}$ model.

We now delve into the basics of head-eye coordination (readers familiar with the basic concepts may skip to Section 2.2), followed by a review and analysis of related work on gaze animation.

2.1 Head-eye Coordination

While previous research [12, 17] has recognized the substantial variability in gaze-driven head motion, there are well-established patterns of head-eye contribution to gaze in psychology literature, that can serve as insights for designing head-eye coordination algorithms.

Ocular Motor Range (OMR) refers to the mechanical limit of eye rotations (a maximum of 45° horizontally and 30° vertically) [11]. In practical scenarios, the eyes seldom reach this limit [10], as the head turns to prevent eye strain. The head’s mobility is confined by the **Cervical Range of Motion**, (a maximum of 90° in the yaw axis, 80° of extension (facing up), 50° of flexion (facing down) in the pitch axis [18]. The degree of head contribution is determined mainly by the amplitude of gaze shift, which we will denote as the **amplitude effect**. For small gaze shifts ($0^\circ - 20^\circ$), predominantly the eyes move. For larger gaze shifts ($20^\circ - 90^\circ$), both the eye and head move, with the amount of head movement positively correlated with gaze amplitude [9, 12]. The head also shows a **midline effect**, suggesting larger and faster head movements when the gaze shift moves the head towards the torso midline [19]. Individual preferences, known as **head propensity**, also play a role in head motion [13]: some people are head movers, who always turn to face gaze targets, while others only turn their heads when necessary. The **dwell time** on a gaze target also determines the degree of head rotation [14], with a greater head turn for a longer intended dwell. They also found that if we expect to continue a second gaze shift after an initial gaze shift, the first gaze shift would have a greater head turn, which we refer to as the **expectation effect** [14].

A number of factors including image features of visual stimuli affect reaction time/saccadic latency [20]. Many factors influence the reaction time (saccadic latency) [20] and the delay between head and eye movements. Typically the head lags behind the eyes slightly [12]. The head moves first for audio triggered gaze shifts [21], and head and hand motion onset is synchronized for tasks involving hands [22].

2.2 Related Work on Gaze Control

While substantial research focuses on the psychology and neural mechanism of gaze control [7], relatively fewer approaches exist that actually generate head-eye coordinated motion, addressing limited subsets of psychological insights (summarized in Table 1). Research focused on saccadic gaze shifts typically uses a common framework based on

Table 1: Comparison of psychological insights addressed by prior art

	Amplitude Effect	Midline Effect	Head Propensity	Dwell Time	Expectation Effect
Itti [23]	✓	✓	✗	✗	✗
Eyecatch [24]	✓	✗	✗	✗	✗
Andrist [19]	✓	✓	✓	✗	✗
Pejsa [25]	✓	✓	✓	✗	✗
Jin [26]	✓	✗	✗	✗	✗
Klein [27]	✗	✗	✗	✗	✗
Goude [28]	✓	✗	✗	✗	✗
Proposed	✓	✓	✓	✓	✓

Motion Summation [19, 23–25, 28–31]. In this framework, the input is modeled by a gaze sequence $\{\mathbf{g}_n, t_n\}_{n=1}^N$ of discrete gaze target positions \mathbf{g}_n , and times they are observed t_n . The output comprises head and eye motion trajectories $\mathbf{h}(t)$, $\mathbf{e}(t)$, such that the combined head-eye movement meets the input gaze targets at the specified times. The sparse input representation aligns with the saccade control circuit proposed in psychological literature [32], and makes animator authoring intuitive [30]. First, a sequence of head-eye configurations that satisfy the input gaze targets is generated. Subsequently, these head and eye configurations are interpolated by using a discrete integrator along with gaze shift velocity and duration parameters, to output head and eye motion trajectories (see Section 4.1).

The research under this framework primarily differs in their approach to computing the head and eye configurations that satisfy the input gaze targets, the velocity profiles used, and duration of each head-eye shift (see video for visual comparisons).

The head’s contribution to gaze shifts is observably important: the **amplitude effect** [9] has been used to compute head contribution [24, 28], where the head only moves towards the gaze target if the angle between the current head orientation and the target exceeds specific thresholds (20° for [24], 40° for [28]). Itti et al. [23] threshold the head turning towards gaze targets that move the head towards the midline (**midline effect**). Previous works [19, 25, 30] have employed the midline effect, and further modulated the head contribution by a user-specified head turning propensity (ranging from minimal head turn needed once the eye is at its mechanical limit, to maximal head turn towards target).

The head velocity profile and movement duration have a subtle but noticeable effect on the perceptual quality of the animation. [19, 25, 30] use a piece-wise polynomial velocity profile and very short duration for the head shift. While convincing for a single gaze shift, the motion is perceptually choppy for a sequence of closely timed gaze shifts. [28] linearly accelerate/decelerate the head between 0 and $40^\circ/sec$, resulting in a smooth, yet sluggish appearance for quick gaze shifts. In general, the quadratic velocity profile used by [23] and the minimal jerk velocity profile [24] better match our empirical observations.

Gaze shifts exhibit more irregular velocity profiles than head movements [33], with dynamics that vary based on task context—for instance, gaze shifts during hand-eye coordination are faster than those for visual observation alone [34]. However, given that gaze shifts are extremely brief (20-40ms), these velocity profile differences become imperceptible to observers. We therefore adopt a simplified ease-in-ease-out velocity profile used in existing animation systems [23, 24].

Data-driven methods for head-eye coordination also exist. [27] propose a recurrent neural network (RNN) to learn head and eye motion dynamics from an in-house dataset of actors performing smooth pursuit gaze tasks. While good at reproducing smooth pursuit, saccadic motion is unresponsively smooth. Data-driven methods have also been used to generate expressive, emotional, or stylized gaze transitions [35, 36]. Speech driven models for head and eye motion that both, predict gaze shifts and generate their head and eye motion trajectories [37] [38] and [26] (LSTM), tend to be context specific to conversational gaze patterns. In contrast, our approach complements speech-driven models that decouple conversational gaze sequence prediction, from the head-eye coordination needed to satisfy the gaze shifts (see Section 5:17-5:44 in video). Finally, our work can also benefit research [39, 40]

that produce visual scan paths in images or scenes, but do not generate head or eye motion trajectories.

None of the existing literature accounts for dwell time and the expectation effect (Table 1), which is not surprising since the examination of head-eye coordination in psychology literature predominantly centers on individual gaze shifts [9, 12, 13], that lack quantifiable behavioural patterns for dwell time and expectation. Inspired by [41, 42] that establish the validity of studying gaze in VR, we propose a collection protocol and head-eye coordination model that specifically accounts for intended dwell time and expectation.

3 VR DATA COLLECTION

Traditionally, gaze stimuli have been presented as arrays of LED lights arranged in semicircles or hemispheres around subjects, with gaze tracked using head-mounted pupil and head motion sensors.

Dynamically adjusting the space-time presentation of stimuli based on the subject’s gaze, is difficult with a physical setup, such as making a stimulus disappear after an intended dwell time, or presenting expected stimuli based on stimuli being actively observed.

Findings from [41] demonstrated that gaze shift behavior in VR reflects that observed in physical environments. VR headsets also combine 3D sensing and display, enabling the creation of interactive experiments that dynamically adapt the presentation of gaze targets in response to the subject’s movements. We thus develop a VR framework to study dwell time and expectation effect, and gaze scenarios that contain interactions between dwell time and expected targets.

3.1 Experiment - Dwell Time

[14] show that humans tend to turn their heads towards a gaze target less if they intend to dwell on the target for a short time, and vice-versa. However, it was unclear what constitutes a short dwell, and whether the relationship between dwell time and head turn is continuous (e.g., linear) or discrete (short and long). We thus aim to first replicate the results from [14] to validate our data collection in VR, then examine the relationship between dwell time and the corresponding head contribution.

We thus conducted a single-factor experiment where we presented fixed targets at variable dwell times, and captured the head rotation amplitude Δh at the end of each gaze shift. The experimental conditions chosen based on pilot testing¹ were *target-angle* (20° , -40° , negative looks left), and *dwell-time* (0.035, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0, 1.5 seconds).

The angles (20° and 40°) were chosen so that the target would always be visible in the periphery, enabling the subject to anticipate both the dwell time and target position. For each combination of target-angle and dwell-time, we ran 5 repetitions, i.e. 100 ($2 \times 10 \times 5$) trials per participant. The order of gaze target presentation is randomized during data collection.

To ensure each trial is independent, we start each trial with an initial stimulus (1 second dwell time) directly in front of the subject to reset the head and eyes, and then present the actual stimulus with varying dwell time (figure 2c). This method is widely used in gaze literature [14, 41, 43, 44]. We randomized the order of stimuli presentation across different dwell times and angles to prevent subjects from memorizing the patterns and adapting their behavior through repetition.

3.1.1 Stimulus Design

Unlike [14], where each gaze stimulus appears for a set time (0.5s for short dwell and 1.5s for long dwell), real-time gaze tracking allows us to spatio-temporally adapt the display based on the subject’s gaze.

Figure 2 shows our proposed interaction. Each target (blue) is inactive when it first appears, surrounded by a red circle of variable radius, where the radius indicates the required dwell time (Figure 2a). When

¹We conducted a pilot study ($n=6$) to tested left-right symmetry by recording head amplitude for stimuli at six angle pairs ($\pm 10^\circ$ to $\pm 60^\circ$). A one-sample t-test on mean head movement differences showed no significant directional bias ($t(35) = 0.33$, $p = .747$, $d = 0.055$), justifying our focus on one side per angle.

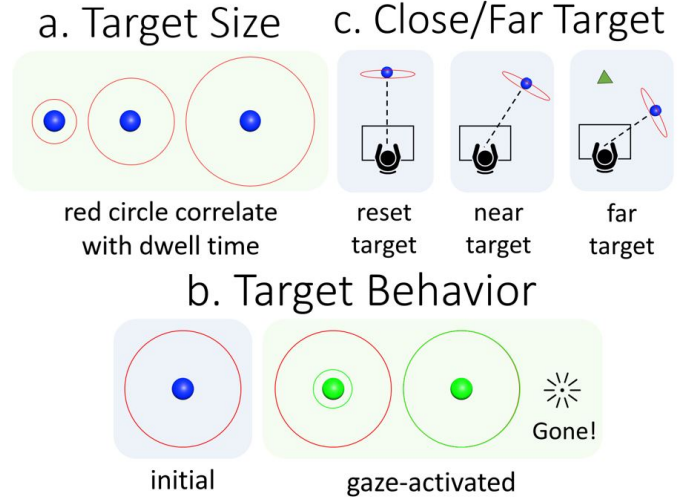


Fig. 2: Gaze target stimulus behavior: (a) The outer circle size informs the subject of the required dwell time. (b) Gazing at the target activates it, causing indicator circle to expand (c) Between each stimulus presentation, a central target resets the subject’s gaze. Distant targets are indicated by green arrows.

the participant’s gaze intersects the target, it turns green and a green indicator circle expands at constant speed until reaching the red boundary, at which point the target disappears and the next appears (Figure 2b). The relative sizes of the red and expanding green circles provide visual feedback on the remaining dwell time. While the expanding animation may influence target salience, this effect is consistent across all experimental conditions and does not confound our results.

In cases where the next target appears outside the subject’s field of view, such as in Section 3.4, an arrow pointing towards the next target is displayed in place of the disappearing target (Figure 2c).

3.1.2 Procedure

The experiment commenced with the subject seated in a non-swivel chair, both arms resting on a table in front of them. Subjects were instructed to keep their elbows on the table to maintain torso stability while allowing free movement of the head. The subjects were briefed on the significance of the red circle’s size as an indicator of the intended dwell time for each target. Then each subject was directed to put on the VR headset, and the calibration software was executed to ensure accurate gaze tracking. The experiment proceeded with the display of 100 trials consecutively. On average, each subject completed the procedure in approximately 6 minutes. All experiments took place at the University of Toronto, approved by the University of Toronto Research Ethics Board under Protocol 38139.

3.1.3 Results

We conducted this experiment with 16 subjects (10 males, 6 females, ages 37.3 ± 17.4). We recorded the presentation time of each gaze stimulus, as well as the head and eye trajectory at 50Hz to avoid non-uniform sampling due to potential frame drop. We discretized the head trajectory into distinct head shifts based on the presentation time of each stimulus. We then obtained the head rotation amplitude Δh by taking the difference in head angle when the stimulus is first presented and when the stimulus disappears. We collected a total of 800 pairs of $\{dwell, \Delta h\}$, in which we observed only 1 gaze-shift towards a non-target position.

We validated our approach by reproducing the analysis from [14]. We used a two-way TARGET AMPLITUDE \times DWELL TIME Repeated Measures ANOVA ($\alpha=.05$). We categorized Dwell Time as *short* ($< 0.5s$), or *long* ($\geq 0.5s$) to study the effect on head motion towards gaze targets. We tested the normality assumption with the Shapiro-Wilk

test. Bonferroni-corrected post hoc tests were used when applicable. The effect sizes are reported as partial eta squared (η_p^2).

Both DWELL TIME ($F_{1,14}=36.73$, $p<.001$, $\eta_p^2=.72$) and TARGET AMPLITUDE ($F_{1,14}=34.69$, $p<.001$, $\eta_p^2=.71$) showed significant main effects on head rotation. We also found a significant TARGET AMPLITUDE \times DWELL TIME interaction ($F_{1,14}=6.64$, $p<.05$, $\eta_p^2=.32$). Post hoc tests (Table 2) showed that DWELL TIME has a significant effect on head movement at both TARGET AMPLITUDES (both $p<.001$). Similarly, TARGET AMPLITUDE also had a significant effect in both DWELL TIMES (both $p<.001$).

Table 2: Impact of long and short dwell time at different target amplitudes.

target angle	short dwell θ_{head}	long dwell θ_{head}
20°	5.8 ± 5.7	10.6 ± 6.5
40°	20.7 ± 9.4	26.5 ± 10.1

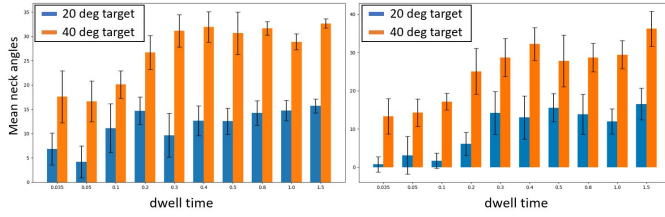


Fig. 3: Two examples of subject head rotation corresponding to stimulus dwell time. The columns represents the head contribution towards the 40°(orange) and 20°(blue) gaze target.

Figure 3 shows two examples of subject head rotation plotted against the dwell time of each stimulus. We see a consistent rise in head contribution with increasing dwell time until it reaches 0.3/0.4s, beyond which the head contribution stabilizes. This reflects the need for our head IK system to employ a continuous parameter reflecting dwell time’s impact on head contribution, but with a saturation when dwell time exceeds 0.3/0.4s. Based on these observations and our results, we can confirm that dwell time has a significant impact on the amount of head rotation. For subsequent experiments, we consider a dwell time less than 0.3s as an unambiguous **short dwell** and a dwell time greater than 0.5s as a **long dwell**.

3.2 Experiment - Expectation Effect

[14] showed that when a subject expects to make two gaze shifts in the same direction, the head movement for the initial gaze shift will have a larger amplitude than usual. One can observe a smaller amplitude when the gaze shifts in the opposite direction, like when watching tennis (see 00:38-00:43 in video). We thus generalize our study to analyze the effect of subsequent gaze targets in the same (*onward condition*) or opposite (*tennis condition*) direction of motion, on the initial head turn.

In the onward condition, we replicated the experimental setup used in the dwell time experiment. Each trial comprised three stimuli (Figure 4a): first, a head-eye reset target with a 1s dwell; second, a shorter 0.1s or 0.5s appeared, accompanied by a hint indicating the position for a subsequent target; third, the secondary target, with a 1.5s dwell appearing in place of the hint. The trial concluded once the subject attended to all three targets. We experimented with varying combinations of primary (20°, -40°) and secondary (none, 10°, 20°, 30°, 40° beyond the primary) target angles. The primary target was presented with both short (0.1s) and long (0.5s) dwell time. Each combination was repeated 5 times, i.e. a total of 100 trials ($2 \times 5 \times 2 \times 5$), presented in randomized order.

In the tennis condition, stimuli are presented as an alternating sequence across the torso mid-line of (Figure 4b). Each trial comprises 11 stimuli: the initial reset stimulus followed by 5 repetitions of targets that alternate left and right at fixed angles and dwell time. We use the

a. Onward condition b. Tennis condition

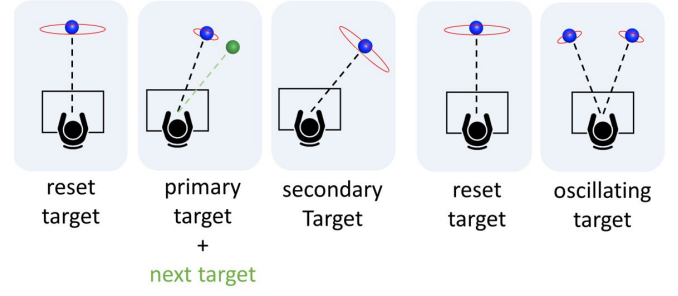


Fig. 4: Expectation effect trial showing onward (a) and tennis (b) condition.

following combinations of target angle (20°, 40°) and dwell time (0.1 for short dwell and 0.5 for long dwell).

Procedure Similar to the dwell time experiment, subjects are instructed to avoid moving their torso but may move their head freely. In both onward and tennis condition testing, subjects can take a break if needed, and are informed of the stimuli pattern to aid anticipatory behavior. Data collection takes approximately 10 minutes.

Results We conducted the onward and tennis experiment on 13 subjects who were also part of the dwell time experiment (8 males, 5 females, ages 33.0 ± 13.1). For the onward condition, we first attempted to replicate the result from [14]. We analyzed Δh (obtained as in Section 3.1) using a TARGET AMPLITUDE \times DWELL TIME \times SECONDARY three-way repeated measures ANOVA using the same analysis methodology as in Section 3.1.3. The results can be found in Table 3. The analysis showed significant main effects for all of TARGET AMPLITUDE ($F_{1,12}=72.06$, $p<.001$, $\eta_p^2=.86$), DWELL TIME ($F_{1,12}=143.72$, $p<.001$, $\eta_p^2=.97$), and SECONDARY ($F_{1,12}=13.95$, $p<.05$, $\eta_p^2=.52$). The results confirm the impact of the secondary target on head rotation at both short and long dwell times for the primary target at various angles, as predicted by [14].

Table 3: Result of the onward condition study, showing average head rotation for attending to different gaze TARGET AMPLITUDE (20° vs 40°), considering the presence of SECONDARY targets (with vs without onward) and intended DWELL TIME (long vs short).

target	dwell	θ_{head} without onward	θ_{head} with onward
20°	Short	3.6 ± 3.7	6.1 ± 4.2
40°	Short	16.3 ± 5.9	19.8 ± 6.9
20°	Long	9.9 ± 6.2	14.2 ± 3.9
40°	Long	26.3 ± 8.7	32.7 ± 6.1

Furthermore, these results further validated our experimental approach and emphasized the need to account for anticipated gaze targets, in computing head rotation for gaze control. Note that we found no meaningful Pearson correlation between the amplitude of the secondary gaze shift and head turn for the primary gaze target (weak linear correlation for long dwell and none for short dwell time).

We also qualitatively observed that subjects tend to merge two head shifts into one, when the primary and secondary targets are near each other. The head angle trajectories in Figure 5 for example, show only one stable head angle for a 10°onward shift (left), instead of two stable head angles for a 40°onward shift (right). This suggests that multiple anticipated and proximal gaze targets can be captured by a single head shift when computing head rotation for gaze control.

For the tennis conditions, we conducted a three-way repeated measures ANOVA with TARGET AMPLITUDE, DWELL TIME, and SECONDARY as factors, consistent with the analysis used for the onward condition. The results did not show significant interactions. However, we found significant TARGET AMPLITUDE ($F_{1,7}=29.98$, $p<.001$,

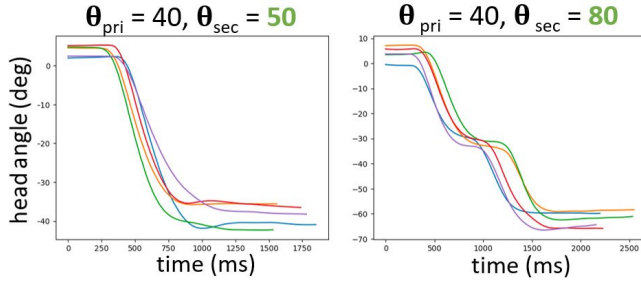


Fig. 5: Head angle trajectories for two different onward condition secondary target angles (5 repetitions, long dwell, same subject).

Table 4: Result of the tennis condition study, showing average head rotation for attending to different gaze TARGET AMPLITUDE (20° vs 40°), considering the presence of SECONDARY targets (with vs without tennis condition) and intended DWELL TIME (long vs short).

target	dwell	θ_{head} without tennis	θ_{head} with tennis
20°	Short	6.1 ± 6.9	9.4 ± 5.4
40°	Short	21.9 ± 9.9	21.1 ± 8.5
20°	Long	10.2 ± 6.8	13.9 ± 5.4
40°	Long	25.9 ± 9.8	25.1 ± 9.6

$\eta_p^2=.81$), and DWELL TIME ($F_{1,7}=11.82$, $p<.05$, $\eta_p^2=.63$) main effects, confirming that intended dwell time remains a significant factor in successive sequential gaze shifts. This effect was not demonstrated in [14], which focused exclusively on single-instance gaze shifts. However, the presence of SECONDARY targets in the opposite direction did not have a significant effect on the degree of head involvement ($F_{1,7}=1.02$, $p>.05$, $\eta_p^2=.13$).

3.3 Key Insights

Expanding on the findings of [14], we discovered a continuous positive correlation between intended dwell time and head contribution to gaze (plateauing for dwell times $> 0.4s$). Our onward expectation effect experiment confirmed energy-conserving head-turn behavior, accounting for anticipated gaze targets. Multiple proximal gaze sequence targets were also satisfied using a single head movement. Our experiment under tennis conditions reinforced that the dwell time effect extends to successive gaze shifts. These insights are critical to the design of our *Head-EyeK* algorithm in Section 4, which can determine the head trajectory when attending to a gaze sequence of arbitrary number of gaze targets.

3.4 Additional Data Collection

While the dwell time and expectation effect experiments allow us to investigate gaze parameters in a controlled fashion, they do not reflect the spontaneous nature of gaze behavior in the wild. To evaluate Head-EyeK, we collected additional motion data within the proposed experimental setup, as well as in augmented reality (AR) and VR gaming sessions, to better capture head-eye coordination behavior in less controlled scenarios. Note that while existing datasets [45] capture head-eye coordination behavior, we opted to create our own dataset because [45] allows free torso movement, which affects head-eye coordination patterns [41]. In the proposed experimental setup, we recorded challenging gaze behaviors, such as the "double take", where the subject first takes a short glance at a target, looks away, then quickly looks back at the target for a second and longer time. We also recorded randomly presented gaze target sequences incorporating 20 randomly ordered instances of isolated gaze shifts, double takes, tennis conditions, and onward conditions.

In AR, we collected head and gaze data from seated users wearing a VR headset in pass-through mode while performing a physical block-stacking task. The participants reconstructed designs illustrated in two reference images (Figure 6 middle). Lastly, we recorded seated

users playing a VR wack-a-mole game, where users smack moles that randomly appeared in a semi-circle in front of them (Figure 6 right).

We collected data from 16 consenting individuals (10 males, 6 females, ages 37.3 ± 17.4), comprising a total of 88 minutes of head-eye coordination, used for evaluation in section 5.

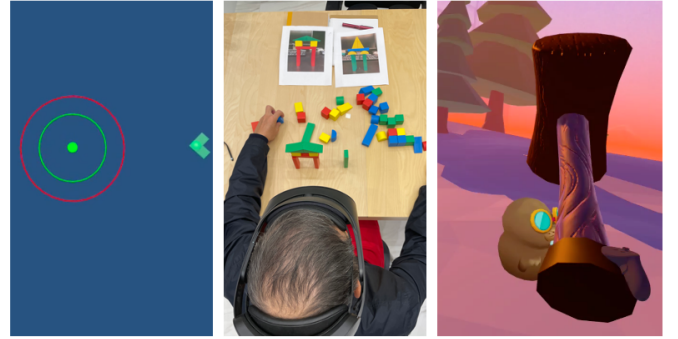


Fig. 6: Additional data collection: controlled gaze target sequences (left), physical block stacking (middle), virtual wack-a-mole game (right).

3.5 Implementation Details

All experiments were conducted on a Meta Quest Pro (a commercial VR headset with head and eye tracking ability) [46]. The headset display provides a 106°(horizontal) by 95°(vertical) field of view, with a refresh rate of 72-90 Hz. Although the field of view (FOV) meets the mechanical limit, the peripheral vision may remain incomplete at the extreme eye ranges. The headset uses 5 infrared cameras for pupil tracking, capable of capturing gaze shifts of up to 60°at up to 90Hz, with a tracking error of 0.85°during head-free gaze shifts [46]. The device combines internal accelerometers with external cameras, to track the head with an average error of 0.48°at up to 90Hz [47]. Our experimental framework, to be publicly released, was developed in Unity.

4 HEAD-EYEK ALGORITHM

We take in a sequence of gaze targets and look-at times $\{\mathbf{g}_n, t_n\}_{n=1}^N$, compute head and eye shifts, then sum up all the motions using an integrator to output head and eye trajectories $\mathbf{h}(t), \mathbf{e}(t)$. The motion summation formulation for head trajectory is:

$$\mathbf{h}(t) = \sum_{n=1}^N \mathbf{h}_i(t) = \sum_{n=1}^N \mathbf{b}_i \cdot v \left(t_i^0, t_i^f, t \right) \quad (1)$$

where each head/eye shift is characterized by: 1) a direction vector \mathbf{b} that determines the direction and magnitude of head/eye shift, 2) the start and end time of the motion t_i^0 and t_i^f , and 3) a velocity profile $v(t_0, t_f, t)$ which specifies the displacement at each time step t during the motion.

Note, we generate gaze trajectory $\mathbf{g}(t)$ similar to [24], then obtain $\mathbf{e}(t) = \mathbf{g}(t) - \mathbf{h}(t)$, as it eliminates the need to account for the Vestibular Ocular Reflex (VOR), and is aligned with gaze being explicitly modeled by neural circuits [48].

Algorithm 1 outlines the overall framework for computing head and eye trajectories. The components of each head shift are computed as follows:

Motion direction vector (\mathbf{b}_n) The motion direction defines the direction and amplitude the gaze/head has to travel during each gaze shift. For gaze, we take it as the difference between the previous gaze target and the current target $\mathbf{b}_{n,gaze} = \mathbf{g}_n - \mathbf{g}_{n-1}$. For head angle, we first compute head gaze targets $\{\mathbf{h}_n, t_n\}_{n=1}^N$ based on minimizing head-eye coordination energy detailed in 4.3, then each head shift direction can be computed similarly as gaze shift directions with $\mathbf{b}_{n,head} = \mathbf{h}_n - \mathbf{h}_{n-1}$.

ALGORITHM 1: Generate Gaze Trajectories

Input: Gaze sequence $\{\mathbf{g}_n, t_n\}_{n=1}^N$
Output: Gaze trajectory $\mathbf{g}(t)$, Head trajectory $\mathbf{h}(t)$
Procedure
 Generate head sequence $\{\mathbf{h}_n, t_n\}_{n=1}^N$ as per Section 4.3;
 Initialize gaze movement set $\mathbb{G} = \{\emptyset\}$;
 Initialize head movement set $\mathbb{H} = \{\emptyset\}$;
 for $n = 1$ **to** N **do**
 generate gaze shift: $\mathbf{b}_{n,gaze}, t_{n,gaze}^0, t_{n,gaze}^f$;
 generate head shift: $\mathbf{b}_{n,head}, t_{n,head}^0, t_{n,head}^f$;
 end
 $\mathbf{g}(t) = \sum_{n=1}^N \mathbf{b}_{n,gaze} \cdot v(t_{n,gaze}^0, t_{n,gaze}^f, t)$;
 $\mathbf{h}(t) = \sum_{n=1}^N \mathbf{b}_{n,head} \cdot v(t_{n,head}^0, t_{n,head}^f, t)$;

Start time (t^0) The start time of each gaze shift are defined by the input gaze target sequence, $t_{n,gaze}^0 = t_n$. For the start time of each head shift, we take $t_{n,head}^0 = t_{n,gaze}^0 + \text{delay}$, reflecting the common observation that the head movement often follows the eye [25]. We set the head delay to have a default value of 100ms.

End time (t^f) The end time of the movement is computed with $t_f = t_0 + \text{duration}$. The movement duration for both gaze and head shifts follows well-studied relationships. The duration of the gaze movement is found to be linearly correlated to the amplitude of the gaze shift [49] (the units are milliseconds and degrees):

$$\text{Gaze duration} = 20 + \|b\|_2 \times 1.33 \quad (2)$$

A similar relation is observed for head shifts, where the peak velocity is linearly correlated to the head shift amplitude [50], which suggests head shifts have a constant duration. We have therefore chosen the value of 400ms, the same as [24], as we have empirically found it to generate the most realistic head motion.

Velocity Profile ($v(t^0, t^f, t)$) chose to use a minimal jerk velocity profile, as it reflects the ease-in ease-out behaviour of natural movement, and it has been empirically shown to reflect natural human motion in [24].

$$v = \frac{30}{(t_f - t^0)^5} \cdot (t - t^0)^2 \cdot (t - t^f)^2 \quad (3)$$

The last step of algorithm 1 generates $\mathbf{g}(t)$ and $\mathbf{h}(t)$. These are converted into trajectory $\mathbf{g}(t)$ and $\mathbf{h}(t)$ by summing/integrating these discrete velocities. In the next section, we discuss how our energy minimization approach generates $\{\mathbf{h}_n, t_n\}_{n=1}^N$.

4.1 Optimization-based Head-Eye Coordination

Our experiments provide various insights about head-eye coordination during a sequence of gaze shifts. However, while our data collection is diverse, it still represents a subset of gaze behavior. Head-eye coordination in gaze involves various cognitive, conversational, and cultural—factors beyond the insights presented in Section 3. Our algorithm is thus designed to provide a procedural structure within which the insights from Section 3 and other factors can be cast as energy terms and optimized.

From the insights obtained from our experiments, we've identified two forms of energy that are minimized during head movements, which we label as **eye strain energy** and **locomotion energy**. In this section, we first discuss how we formulate an energy minimization problem based on these opposing energies, then present our proposed algorithm to solve it and generate $\{\mathbf{h}_n, t_n\}_{n=1}^N$.

eye strain energy Previous works have shown that in head-free gaze shifts, the head always moves such that the eye configuration would not have to be near the Ocular Motor limit [9]. This aligns with our intuition that maintaining our eyes in a rotated position can be

straining. Hence, the natural inclination to turn our heads towards a gaze target can be seen as a mechanism to reduce the energy associated with eye strain.

locomotion energy Inherently, head movements tend to display a sense of economy; unnecessary movements are typically avoided. In shorter dwell times (spanning from 0.1 to 0.4s as in Section 3), the head tends to exhibit minimal movement towards the gaze target, favouring the continuity of the original path or progressing towards the next anticipated target. Conversely, with longer dwell times (0.5s or more), the emphasis on minimizing locomotion becomes less pronounced. Instead, the head would turn towards the gaze target to alleviate eye strain.

Inspired by the observations of these two opposing energies, we have formulated the selection of head angles as an optimization task to minimize the following energy:

$$\{\mathbf{h}_n\}_{n=1}^N = \arg \min_{\{\mathbf{h}_n\}_{n=1}^N} \sum_{n=1}^N \mathbf{E}_{\text{target}}(\mathbf{h}_n) + \mathbf{E}_{\text{transition}}(\mathbf{h}_n, \mathbf{h}_{n+1}) \quad (4)$$

The energy encapsulates two components: the goal of focusing on the gaze target $\mathbf{E}_{\text{target}}$ and managing transitions between different neck configurations $\mathbf{E}_{\text{transition}}$. Both components address one or both of the proposed energies. The target energy is defined as:

$$\begin{aligned} \mathbf{E}_{\text{target}}(\mathbf{h}_n) = & (1 - w(t_n, \text{dwell})) \|\mathbf{h}_n - \mathbf{g}_n\|^2 \\ & + w(t_n, \text{dwell}) \|\mathbf{h}_n - \mathbf{h}_n^*\|^2 + k_{\text{centre}} \cdot \|\mathbf{h}_n\|^2 \end{aligned} \quad (5)$$

This component has three terms respectively. The first of which is the distance between the head direction \mathbf{h}_n and gaze direction \mathbf{g}_n . Directly minimizing this term would completely turn the head to the gaze target, relieving any eye strains. The second term computes the distance between the head direction \mathbf{h}_n to a "lazy" head direction \mathbf{h}_n^* , which is a smoothed version of the gaze-targets (computed by iterative Laplacian smoothing on the gaze-targets, shown in Figure 7). \mathbf{h}_n^* captures an efficient head trajectory that minimizes movement by accounting for both preceding and subsequent gaze targets. Directly minimizing this term would result in a lazy movement of the head that turns towards each target minimally. The first and second terms are weighed by dynamic weighting factors conditioned on dwell time defined as follows:

$$w(t_{\text{dwell}}) = \exp(-k_{\text{dwell}} \cdot t_{\text{dwell}}) \quad (6)$$

$w(t_{\text{dwell}})$ amplifies the importance of conserving locomotion energy when the dwell time is short and emphasizes preventing eye strain when the dwell time is long. The constant k_{dwell} regulates the impact of dwell time, we set it at a default value of 0.4 after conducting a line search to align with the observed effects in the dwell time condition experiment. The last term is the magnitude of \mathbf{h}_n , which reflects the center bias observed in head shifts [9].

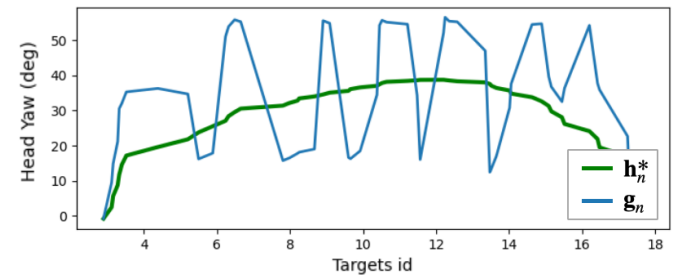


Fig. 7: Comparing head angles when turned towards the "lazy" head direction \mathbf{h}_n^* vs fully turning to the gaze target \mathbf{g}_n

The transition energy is defined as follows:

$$\mathbf{E}_{\text{transition}}(\mathbf{h}_n, \mathbf{h}_{n-1}) = k_{\text{transition}} \cdot w(t_n, \text{dwell}) \|\mathbf{h}_n - \mathbf{h}_{n-1}\|^2 \quad (7)$$

This term gauges the distance necessary to transition between previous and current head angles, approximating the required locomotion effort. Similar to E_{target} , the impact of this term varies with dwell time, being more influential when dwell time is shorter to represent the reduced head locomotion observed in the dwell time experiment. We additionally weigh this term with a customizable constant $k_{transition}$, which we set with a default value of 0.6, obtained through a grid search to minimize the mean square error to the trajectories captured in section 3.4.

We solve the optimization problem by building graphs that represent the proposed energy and solve for the shortest path. We start with the construction of the graphs. Our approach involves two separate graphs—one for yaw (left-right) and another for pitch (up-down) head rotations. Despite their distinct dimensions, both graphs share parallel principles and structures. For simplicity, we’ll focus our discussion on one graph, as both follow the same processes.

The nodes of these graphs are labelled with the convention (frame, angle), summing up to a total of $90 * N + 2$ nodes as shown in Figure 8, where N is the total number of gaze targets to attend to. Each node represents a potential head angle at a certain frame. The angles cover the discretized range of head motion, spanning from -90 to 90 degrees, divided into 2-degree intervals, while the frames range from 1 to N , corresponding to the number of gaze targets in the input gaze sequence. Nodes of consecutive frames are connected via directed edges.

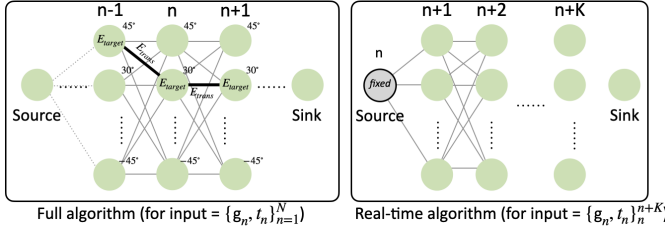


Fig. 8: Graph construction for the proposed method showing node and edge value for nodes $(n-1, 90)$, $(n, 88)$ and $(n+1, 88)$ (left), graph construction for the real-time version of the proposed method, taking K next targets as input, with $K \in \{1, 2, 3, \dots\}$ (right)

The edges encapsulate the transition energy $E_{transition}$ between the two nodes it connects. The nodes themselves are annotated with the target energy E_{target} . Lastly, we add a source node (connected to all nodes in layer 1 with an edge weight of 0), and a sink node (connected to layer N). To solve the optimization problem, we find the shortest path from the source to the sink node. In our implementation, we solve the shortest path between the source and sink node using Dijkstra’s algorithm. Dijkstra has a complexity of $\mathcal{O}((V + E) \log V)$, in our case, it scales with the number of gaze targets N with the relationship $\mathcal{O}(N \log N)$, which is tractable even for large number of gaze targets. The resultant path would then be the head target sequence $\{\mathbf{h}_n, t_n\}_{n=1}^N$. Note that since the source and sink are connected to the rest of the graph with zero-weight edges, they do not influence the overall path.

4.2 Real-time Application

While our algorithm is designed for offline animation, we also proposed a windowed approach for more real-time applications. In the real-time version, instead of constructing a graph based on the entire gaze sequence $\{\mathbf{g}_n, t_n\}_{n=1}^N$, at each time step n , we construct a sub-graph using only the next K gaze targets, $\{\mathbf{g}_n, t_n\}_{n=n+K}^{n+K}$. Different from the full method, instead of having an arbitrary source node, the windowed approach fixes the source node to be the result from the previous window to ensure the motion is continuous. The node is connected to the first of the graph with edges with value $E_{transition}$ like all other nodes. We evaluated the windowing approach against the full approach by comparing the Mean Square Error (MSE) with the ground truth head angle contribution. We found that when only a few look-at-targets are considered, the windowed approach is a crude approximation of the full algorithm. However, with $K \geq 3$, the windowed approach reaches a

similar performance as the full method. This shows that the anticipation of future targets has a sizeable effect on the head contribution of each gaze shift.

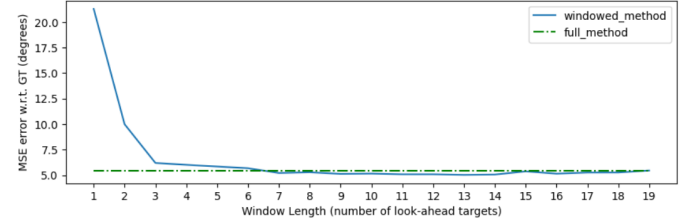


Fig. 9: Comparing ground truth MSE between the windowed approach with the full method on an example gaze sequence

4.3 Smooth Pursuit

Smooth pursuit is a class of actions that describe eye and head motion when continuously tracking a moving target. While our model is not designed for smooth pursuit, many tasks that involve pursuit also contain saccadic movements when switching between pursuit targets, such as looking at different targets during juggling, or switching between rows of text when reading a book.

Instead of a sequence of discrete targets $\{\mathbf{g}_n, t_n\}_{n=1}^N$, to represent moving targets to track, we take two streams of inputs, an object list $\{\mathbf{g}_m(t)\}_{m=1}^M$, which describes the trajectory $\mathbf{g}_m(t)$ of all M objects in the scene, and a pointer list $P(t)$, which determines which of the M objects to track at each time. The sequence $\mathbf{g}_{P(t)}(t)$ would then represent the instantaneous look-at-point. A change in the value of $P(t)$ would indicate that a saccade has occurred, shifting the attention from tracking one object to another. Such a gaze trajectory would alternate between saccadic shifts between different targets, and smooth tracking when there is no target change. Our head-eye optimization model can generate a more realistic head contribution for these saccadic movements between targets.

To compute the optimized head-path for the saccades in this sequence, we consider a sequence of smooth pursuit targets of start position, end position, start and end time $\{\mathbf{g}_n^0, \mathbf{g}_n^1, t_n^0, t_n^1\}_{n=1}^N$, and construct a graph as shown in Figure 8 but with $2N + 2$ nodes since there are now both the start and end targets. For each nodes corresponding to the end angle, we give it the dwell time of $t_n^1 - t_n^0$, as we expect the smooth pursuit to allow the head to catch up to the target, and for the nodes corresponding to the start angle, we consider it having a dwell time of 0, since a combination of saccade and smooth pursuit will be used to reach the gaze target. After solving for the shortest path, the optimized head angles for each saccade would then be the $\{\mathbf{g}_n^{0*}\}_{n=1}^N$.

We integrate smooth pursuit into the saccadic model by modifying algorithm 1. At each timestamp, we specify a additional small displacement towards the gaze target,

$$\dot{\mathbf{h}}_{pursuit}(t) = dt \cdot (\mathbf{g}_{P(t)} - \mathbf{h}(t)) \quad (8)$$

Where $\dot{\mathbf{h}}_{pursuit}(t)$ is the instantaneous velocity of the head and $\mathbf{h}(t)$ is the current head position. To ensure smooth movement, we cap the pursuit velocity at capped at $20^\circ/sec$ for head movement and $100^\circ/sec$ for eye displacement.

5 EVALUATION

Our evaluation is threefold: we show Head-EyeK applied to a sequence of gaze targets defined for reading, juggling and conversation; we then implement and compare 7 state-of-the-art models against ours relative to the 88 minutes of head-eye data collected in Section 3; and we perform a perceptual study to compare Head-EyeK against prior art on 4 gaze tasks.

5.1 Gaze Animation Applications

We choose two complex gaze animation tasks of reading and juggling to showcase Head-EyeK’s ability to optimize for both discrete gaze shifts and smooth pursuit. For reading, we procedurally generate an

input gaze sequence that has segments of smooth pursuit (left to right), and gaze shifts (right to left and down) interleaved (video 4:46-4:53). For juggling, we use balls trajectories tracked from a juggling video (video 4:41-4:46) [51], and an input gaze sequence where gaze shifts to a ball just prior to its apex and smoothly pursues it until a gaze shift to the next ball (as described by jugglers).

Head-eye coordination is also important for expressive conversational gaze. Head-EyeK integrates seamlessly to control the head and eyes of an audio-driven animation system that generates a sparse input sequence of conversational gaze shifts [38] (video 5:16-5:40).

5.2 Prior Art Implementations

For comparison, we have implemented various prior art approaches in Python (available in the supplementary repository). All the procedural models including [29] [19] [23] [24] [19] [25] [28] have been implemented either exactly as specified in the original papers or modified from the provided code. On the other hand, [26] includes a learned prior of head-gaze coordination trained on a 1.4-hour dataset that is no longer available. We therefore trained it based on the 2.25-hour "Gaze in the Wild" dataset [45].

The procedural models expect inputs in the form of a gaze sequence $\{\mathbf{g}_n, t_n\}_{n=1}^N$, which specifies the sequence of gaze targets and the corresponding time each gaze targets are attended to. We obtained these from our collected gaze trajectory $\mathbf{g}(t)$ using the dispersion filter technique with a dispersion threshold of 3° and duration threshold of 0.1 seconds [52], a common method of obtaining intervals of fixation from continuous gaze data. The method involves sliding a window through the trajectory, and calculating dispersion within the window. When the dispersion exceeds the set threshold, it marks the end of a potential fixation. If the duration criteria are met within the window, it identifies the points as fixations and records the centroids \mathbf{g}_n and start times t_n of that fixation. The dispersion of a gaze sequence $\mathbf{g}(t_1, \dots, t_M) = \{x_i, y_i\}_{i=1}^M$ is defined as: $D = \max(x_i) - \min(x_i) + \max(y_i) - \min(y_i)$.

As [26] expect a smooth input gaze trajectory, the gaze data used as is. Both [19] and [25] incorporate an extra scalar parameter, head propensity, to regulate head contribution. To ensure a fair comparison, we found an optimal propensity value by performing a line search to minimize the RMS error across all collected clips. All models are evaluated using the collected gaze trajectory as outlined in Section 3.4, re-sampled to 30 fps. We will compare trajectories for the double-take condition, random condition, wack-a-mole and see-through tasks.

5.3 Quantitative Results

The generated head trajectories are compared quantitatively against ground truth using the average RMSE per frame, shown in Table 5.

Table 5: Average RMSE Comparison with other Baselines

Methods	double look	random	wack-a -mole	see through
Eyecatch [24]	12.10	21.36	46.60	26.10
Goude [28]	21.71	28.63	54.20	24.11
Jin [26]	15.80	29.08	49.90	24.24
Pejsa [25]	9.37	20.09	39.09	24.46
Andrist [19]	9.75	22.66	55.04	25.70
Itti [53]	9.95	20.09	39.19	24.65
Proposed Head-EyeK	8.54	9.47	21.28	13.07

Head-EyeK consistently achieves the lowest error across all conditions compared to prior models. The outcome is unsurprising given that we are unique in considering gaze dwell time and anticipated targets. Models like Eyecatch [24] and Goude [28] trigger head shifts solely based on a gaze shift amplitude threshold without considering relative positions to the previous or future gaze targets. Consequently, when dealing with sequences of small saccades like those in our collected data, these models struggle to generate relevant head rotations. Jin [26] also achieves a high error as the approach generates motion in a Markov fashion, only considering the previous frame of gaze position. This approach generates smooth head motion when given smoothly varying

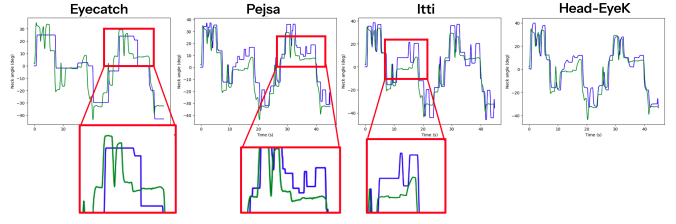


Fig. 10: Head yaw angle for Eyecatch [24], Pejsa [25], Itti [23], and Head-EyeK in the double-take condition. The blue represents the predicted head trajectory, while the green represents ground truth. The figure highlights the lack of head movements in Eyecatch [24], the jittery head movements in Pejsa [30], the exaggerated large movements in Itti [53], and the more reasonable predictions made by Head-EyeK.

gaze trajectories and performs poorly when given gaze trajectories dominated by saccades. The remaining models display relatively reasonable errors across all conditions.

Table 6: Average RMSE Comparison with Ablated Version of Head-EyeK

Methods	double look	random	wack-a -mole	see through
$k_{transition} = 0$	9.30	10.00	21.49	14.98
$w(t_{n,dwell}) = 0$	9.31	10.00	21.53	14.18
$k_{center} = 0$	9.88	10.30	21.41	15.73
Proposed Head-EyeK	8.54	9.47	21.28	13.07

We also conducted an ablation study to evaluate the contribution of each component in our optimization energy function by setting different optimization weights to zero, as shown in Table 6. The results demonstrate that each term contributes to improved head-turn prediction accuracy, validating our design choices. Additionally, since all ablated models maintain reasonable performance, this indicates that our approach is robust to hyperparameter variations, allowing artists to adjust the model parameters without generating infeasible motion.

5.4 Perceptual Study

We animated four challenging gaze behaviors: the double-take and tennis condition (Section 3.3), as well as reading and juggling (Section 5.1) using Head-EyeK and prior art (Table 1). To validate the importance of combining smooth pursuit and discrete gaze shifts, we added a smooth pursuit only animation for the juggling and reading tasks.

We then conducted a perceptual study to rank these animations in their ability to capture the behavior shown in reference videos (see supplemental study videos). We found by pilot testing that a max of 4 animations could be viewed simultaneously and effectively ranked after viewing them freely multiple times. We thus split the 7 (or 8) animations arbitrarily into two sets, both of which included Head-EyeK.

Viewers were thus shown 8 randomized videos (2 split * 4 gaze tasks). We collected responses from 25 viewers (university students) recruited remotely, for the 15-minute study deployed via a web browser. Table 7 shows that Head-EyeK was ranked first or second for all trials, and in particular was an improvement over using smooth pursuit only for juggling and reading. We attribute second rank on double-takes to our reference video being more exaggerated in motion than our input gaze targets, encouraging viewers to prefer the exaggerated and jittery head movements demonstrated by Itti et al. [53], and Jin et al. [26].

We further evaluated a ranking of Head-EyeK and prior approaches using the Plackett-Luce model [54], which estimates the relative "worth" of each method (Table 8). Overall, Head-EyeK scored 4 times better than the next best approach.

To gain a better understanding of why Head-EyeK outperforms prior methods, we present the motion curves of the top four approaches from the user study in Figure 10. As seen, the proposed approach generates a trajectory that is much more closely aligned with the ground truth, resulting in significantly more natural motions.

Table 7: Result for preference study (N=25), showing the overall rank of each method, with rank 1 = most preferred and rank 4 = least preferred.

examples	ranked 1	ranked 2	ranked 3	ranked 4
double take	Jin	Head-EyeK	Goude	Andrist
double take	Itti	Head-EyeK	Yeo	Pejsa
juggling	Head-EyeK	pursuit only	Pejsa	Andrist
juggling	Head-EyeK	Goude	Itti	Yeo
reading	Head-EyeK	pursuit only	Yeo	Pejsa
reading	Head-EyeK	Itti	Jin	Goude
tennis	Head-EyeK	Pejsa	Yeo	Itti
tennis	Head-EyeK	Andrist	Goude	Jin

Table 8: Computed worth, z-values, and p-values for different approaches using the Plackett-Luce model. The z-values represent the standardized differences concerning the proposed method.

Method	Worth	Z-Value	P-Value
Head-EyeK	0.4290	—	—
Eyecatch [24]	0.1022	11.09	<0.01
Itti [23]	0.1377	9.88	<0.01
Pejsa [30]	0.1066	10.94	<0.01
Andrist [55]	0.0709	12.15	<0.01
Goude [28]	0.0795	11.86	<0.01
Jin [26]	0.0741	12.04	<0.01

5.5 Discussion

Summary of contribution: In this paper, we introduced Head-EyeK, a bio-mechanically inspired head-eye coordination model that leverages dwell time and the anticipation of future gaze points to generate accurate head movements during gaze shifts and smooth pursuit across various scenarios. We conducted behavioral psychology experiments, which demonstrated that both dwell time and future gaze anticipation significantly influence head movement. These findings informed the design of Head-EyeK. Our validation is twofold. First, our model quantitatively outperforms prior works when compared with ground truth. Second, an extensive user study showed that Head-EyeK generates more natural motion compared to previous approaches.

Number of Look-Ahead Targets: In Section 4.2, we experimented with varying the number of look-ahead targets to examine its impact on predicted head rotation and whether it influences the mean square error relative to the ground truth, aiming to balance runtime and accuracy. Our findings show that by simply considering two future targets, rather than only the immediate next gaze target, we achieved the most significant performance improvement, with further increases in window size yielding only incremental gains. This underscores the importance of accounting for anticipation in gaze modelling. Additionally, we observed that accounting for three look-ahead targets allowed us to predict head contributions to gaze shifts almost as accurately as when considering all future targets. This result offers potential insights into the neural pathways underlying head-eye coordination, which could be explored in future research.

Focus on Head-Gaze: Accurately predicting head movements based on gaze behavior holds significant potential for improving head-mounted display (HMD) rendering techniques. Similar to foveated rendering, where computational resources are concentrated on the part of the scene the user’s gaze is focused on, head-gaze prediction could enable systems to anticipate where a user’s head will turn next. By pre-rendering or prioritizing the rendering of areas in the user’s peripheral vision that are likely to become the focus, significant computational resources could be saved. This approach would not only reduce the overall computational load but could also enable more efficient, real-time rendering in resource-intensive virtual environments.

Convincing Virtual Characters: Creating realistic and convincing virtual humans requires solutions spanning multiple disciplines,

including graphics, animation, and behavioral psychology. While recent advancements in both software and hardware have significantly improved rendering quality, the increased realism often exposes subtle deficiencies in motion dynamics, particularly in eye and head movements [56]. As the demand for photo-realistic characters rises, it becomes critical to address these motion inconsistencies to avoid the uncanny valley effect, where characters appear unsettling due to unnatural movements. By providing a system capable of generating dynamic and natural head-eye coordination, we aim to bridge this gap, enhancing the believability of virtual humans. Our approach helps ensure that motion quality keeps pace with rendering advancements, leading to more lifelike and engaging virtual characters.

Further Applications: Although our evaluation of Head-EyeK focuses on generating believable avatar behavior, the ability to accurately predict head orientation has significant implications for ergonomics in head-mounted displays. For instance, VR applications with extended screen time, such as virtual workspaces and classrooms, can leverage Head-EyeK’s predictive capabilities to predict head angles based on the user interface (UI) layouts, enabling designers to optimize UI layouts for reducing head strain and enhancing user comfort. Additionally, the adjustable parameters of Head-EyeK— k_{dwell} , k_{center} , and $k_{transition}$ —can be tailored to individual users, generating personalized head motion profiles, which can be used to dynamically adapt UI layouts across various applications, further improving ergonomic design and reducing discomfort during prolonged use.

VR-based Gaze Experiment: Traditionally, studying gaze behavior requires the construction of physical light arrays to present stimuli to subjects [13–15], which can be both tedious and costly to set up. By replicating existing experiments within our VR setup [14], our findings align with prior work [41] and support the validity of using VR as an effective, low-cost alternative to traditional physical setups for studying gaze behavior. With pass-through mode and the potential of using game engine to render realistic virtual environments, the VR medium may also enable studies of in-the-wild behavior of head-eye coordination, which can be explored in future work.

Limitations: While our model generates more accurate head movements compared to previous approaches, it still has several limitations. Firstly, our behavioral studies focus on gaze shifts in the horizontal direction (Sections 3.1 and 3.2), which may limit the generalizability of our findings. Secondly, the head contribution in our model depends on three parameters: k_{dwell} , k_{center} , and $k_{transition}$. Although having multiple parameters enhances expressiveness, adjusting these variables might present challenges for animators, especially when compared to simpler models like [19, 25], which utilize a single “head propensity” parameter for behavior tuning. Lastly, the performance of our model is highly dependent on the quality of the input gaze sequence. If the target gaze sequence is unrealistic or unachievable, the resulting head movements may appear unnatural or undesirable. Lastly, our model is grounded in ergonomic and psychological principles, focusing on natural efficiency. It does not account for more stylized head-gaze behaviors, such as exaggerated side-eye glances, dismissive looks, or eye rolls [35], which are often used to convey specific emotional or stylistic cues. Incorporating such behaviors would require additional layers of customization for expressive purposes.

6 CONCLUSION

In conclusion, we present a new framework for studying human gaze behavior in VR. We illustrate the framework with studies on intended dwell time and gaze target expectation, that validate prior findings, and produce new insights, revealing that head motion planning aligns with principles of energy conservation. We use our findings to formulate a novel Head-EyeK algorithm, that we comprehensively evaluate and show to perform better than prior art. Head-eye coordination for gaze animation is a fundamental aspect of facial animation and digital character behaviour, and we hope our open source framework and implementations of various algorithms will provide a foundation for future work on gaze control and animation.

REFERENCES

- [1] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell, "A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception," *Computer Graphics Forum*, vol. 34, no. 6, pp. 299–326, 2015. 1
- [2] X. Meng, R. Du, and A. Varshney, "Eye-dominance-guided Foveated Rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1972–1980, May 2020. 1
- [3] J. M. Evangelista Belo, A. M. Feit, T. Feuchtnner, and K. Grønbaek, "XR-ergonomics: Facilitating the Creation of Ergonomic 3D Interfaces," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–11. 1
- [4] Y. Zhang, K. Chen, and Q. Sun, "Toward optimized vr/ar ergonomics: Modeling and predicting user neck muscle contraction," in *ACM SIGGRAPH 2023 Conference Proceedings*, ser. SIGGRAPH '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3588432.3591495> 1
- [5] N. Sendhilnathan, T. Zhang, B. Laffreniere, T. Grossman, and T. R. Jonker, "Detecting input recognition errors and user errors using gaze dynamics in virtual reality," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3526113.3545628> 1
- [6] R. Henrikson, T. Grossman, S. Trowbridge, D. Wigdor, and H. Benko, "Head-coupled kinematic template matching: A prediction model for ray pointing in vr," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–14. [Online]. Available: <https://doi.org/10.1145/3313831.3376489> 1
- [7] D. Guitton, "Control of eye-head coordination during orienting gaze shifts," *Trends in Neurosciences*, vol. 15, no. 5, pp. 174–179, May 1992. 1, 2
- [8] A. F. Fuchs, "Saccadic and smooth pursuit eye movements in the monkey," *The Journal of Physiology*, vol. 191, no. 3, pp. 609–631, Aug. 1967. 1
- [9] E. G. Freedman and D. L. Sparks, "Eye-Head Coordination During Head-Unrestrained Gaze Shifts in Rhesus Monkeys," *Journal of Neurophysiology*, vol. 77, no. 5, pp. 2328–2348, May 1997. 1, 2, 3, 6
- [10] A. Frisken, A. P. Bayliss, and S. P. Tipper, "Gaze Cueing of Attention," *Psychological bulletin*, vol. 133, no. 4, pp. 694–724, Jul. 2007. 1, 2
- [11] Y. Shin, H. Lim, M. Kang, M. Seong, H. Cho, and J. Kim, "Normal range of eye movement and its relationship to age," *Acta Ophthalmologica*, vol. 94, no. S256, 2016. 1, 2
- [12] J. B. Pelz and R. Canosa, "Oculomotor behavior and perceptual strategies in complex tasks," *Vision Research*, vol. 41, no. 25–26, pp. 3587–3596, 2001. 1, 2, 3
- [13] J. H. Fuller, "Head movement propensity," *Experimental Brain Research*, vol. 92, no. 1, pp. 152–164, 1992. 1, 2, 3, 9
- [14] B. S. Oommen, R. M. Smith, and J. S. Stahl, "The influence of future gaze orientation upon eye-head coupling during saccades," *Experimental Brain Research*, vol. 155, no. 1, pp. 9–18, Mar. 2004. 1, 2, 3, 4, 5, 9
- [15] E. G. Freedman, "Coordination of the Eyes and Head during Visual Orienting," *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, vol. 190, no. 4, pp. 369–387, Oct. 2008. 2, 9
- [16] J. D. Crawford, J. C. Martinez-Trujillo, and E. M. Klier, "Neural control of three-dimensional eye and head movements," *Current Opinion in Neurobiology*, vol. 13, no. 6, pp. 655–662, Dec. 2003. 2
- [17] L. Borel, B. Le Goff, O. Charade, and A. Berthoz, "Gaze strategies during linear motion in head-free humans," *Journal of Neurophysiology*, vol. 72, no. 5, pp. 2451–2466, Nov. 1994. 2
- [18] E. E. Swartz, R. T. Floyd, and M. Cendoma, "Cervical Spine Functional Anatomy and the Biomechanics of Injury Due to Compressive Loading," *Journal of Athletic Training*, vol. 40, no. 3, pp. 155–161, 2005. 2
- [19] S. Andrist, T. Pejasa, B. Mutlu, and M. Gleicher, "A head-eye coordination model for animating gaze shifts of virtual characters," in *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. Santa Monica California: ACM, Oct. 2012, pp. 1–6. 2, 8, 9
- [20] B. Duinkharjav, P. Chakravarthula, R. Brown, A. Patney, and Q. Sun, "Image features influence reaction time: a learned probabilistic perceptual model for saccade latency," *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 144:1–144:15, Jul. 2022. [Online]. Available: <https://dl.acm.org/doi/10.1145/3528223.3530055> 2
- [21] J. E. Goldring, M. C. Dorris, B. D. Corneil, P. A. Ballantyne, and D. P. Munoz, "Combined eye-head gaze shifts to visual and auditory targets in humans," *Experimental Brain Research*, vol. 111, no. 1, pp. 68–78, Sep. 1996. 2
- [22] J. B. Smeets, M. M. Hayhoe, and D. H. Ballard, "Goal-directed arm movements change eye-head coordination," *Experimental Brain Research*, vol. 109, no. 3, pp. 434–440, Jun. 1996. 2
- [23] L. Itti, N. Dhavale, and F. Pighin, "Photorealistic Attention-Based Gaze Animation," in *2006 IEEE International Conference on Multimedia and Expo*. Toronto, ON, Canada: IEEE, Jul. 2006, pp. 521–524. 2, 8, 9
- [24] S. H. Yeo, M. Lesmana, D. R. Neog, and D. K. Pai, "Eyecatch: Simulating visuomotor coordination for object interception," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 42:1–42:10, Jul. 2012. 2, 5, 6, 8, 9
- [25] T. Pejasa, S. Andrist, M. Gleicher, and B. Mutlu, "Gaze and Attention Management for Embodied Conversational Agents," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 1, pp. 1–34, Mar. 2015. 2, 6, 8, 9
- [26] A. Jin, Q. Deng, Y. Zhang, and Z. Deng, "A Deep Learning-Based Model for Head and Eye Motion Generation in Three-party Conversations," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 2, no. 2, pp. 1–19, Jul. 2019. 2, 8, 9
- [27] A. Klein, Z. Yumak, A. Beij, and A. F. van der Stappen, "Data-driven Gaze Animation using Recurrent Neural Networks," in *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, ser. MIG '19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 1–11. 2
- [28] I. Goudé, A. Bruckert, A.-H. Olivier, J. Pettré, R. Cozot, K. Bouatouch, M. Christie, and L. Hoyet, "Real-time Multi-map Saliency-driven Gaze Behavior for Non-conversational Characters," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–13, 2023. 2, 8, 9
- [29] O. Oyekoya, W. Steptoe, and A. Steed, "A saliency-based method of simulating visual attention in virtual scenes," in *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. Kyoto Japan: ACM, Nov. 2009, pp. 199–206. 2, 8
- [30] T. Pejasa, D. Rakita, B. Mutlu, and M. Gleicher, "Authoring directed gaze for full-body motion capture," *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 1–11, Nov. 2016. 2, 8, 9
- [31] E. Kokkinara, O. Oyekoya, and A. Steed, "Modelling selective visual attention for autonomous virtual characters," vol. 22, no. 4, pp. 361–369. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/cav.425> 2
- [32] C. A. Scudder, C. S. Kaneko, and A. F. Fuchs, "The brainstem burst generator for saccadic eye movements: A modern synthesis," *Experimental Brain Research*, vol. 142, no. 4, pp. 439–462, Feb. 2002. 2
- [33] D. Tweed, B. Glenn, and T. Vilis, "Eye-head coordination during large gaze shifts," vol. 73, no. 2, pp. 766–779. 2
- [34] J. Epelboim, R. M. Steinman, E. Kowler, Z. Pizlo, C. J. Erkelens, and H. Collewijn, "Gaze-shift dynamics in two kinds of sequential looking tasks," *Vision Research*, vol. 37, no. 18, pp. 2597–2607, Sep. 1997. 2
- [35] Y. Ferstl, "Generating emotionally expressive look-at animation," in *Proceedings of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games*, ser. MIG '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3623264.3624438> 2, 9
- [36] X. Ma and Z. Deng, "Natural Eye Motion Synthesis by Modeling Gaze-Head Coupling," in *2009 IEEE Virtual Reality Conference*, Mar. 2009, pp. 143–150, ISSN: 2375-5334. [Online]. Available: <https://ieeexplore.ieee.org/document/4811014> 2
- [37] B. H. Le, X. Ma, and Z. Deng, "Live Speech Driven Head-and-Eye Motion Generators," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 11, pp. 1902–1914, Nov. 2012, conference Name: IEEE Transactions on Visualization and Computer Graphics. [Online]. Available: <https://ieeexplore.ieee.org/document/6165277> 2
- [38] Y. Pan, R. Agrawal, and K. Singh, "S3: Speech, script and scene driven head and eye animation," *ACM Trans. Graph.*, vol. 43, no. 4, Jul. 2024. [Online]. Available: <https://doi.org/10.1145/3658172> 2, 8
- [39] D. Martin, A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia, "ScanGAN360: A Generative Model of Realistic Scanpaths for 360° Images," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2003–2013, May 2022, conference Name: IEEE Transactions on Visualization and Computer Graphics. [Online]. Available: <https://ieeexplore.ieee.org/document/9714046> 2

- [40] G. Boccignone, V. Cuculo, A. D'Amelio, G. Grossi, and R. Lanzarotti, "On Gaze Deployment to Audio-Visual Cues of Social Interactions," *IEEE Access*, vol. 8, pp. 161 630–161 654, 2020. 2
- [41] L. Sidenmark and H. Gellersen, "Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality," *ACM Transactions on Computer-Human Interaction*, vol. 27, no. 1, pp. 4:1–4:40, Dec. 2019. 3, 5, 9
- [42] Z. Hu, A. Bulling, S. Li, and G. Wang, "EHTask: Recognizing User Tasks From Eye and Head Movements in Immersive Virtual Reality," *IEEE transactions on visualization and computer graphics*, vol. 29, no. 4, pp. 1992–2004, Apr. 2023. 3
- [43] H. H. Goossens and A. J. Van Opstal, "Human eye-head coordination in two dimensions under different sensorimotor conditions," *Experimental Brain Research*, vol. 114, no. 3, pp. 542–560, May 1997. 3
- [44] J. S. Stahl, "Amplitude of human head movements associated with horizontal saccades," *Experimental Brain Research*, vol. 126, no. 1, pp. 41–54, May 1999. 3
- [45] R. Kothari, Z. Yang, C. Kanan, R. Bailey, J. B. Pelz, and G. J. Diaz, "Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities," *Scientific Reports*, vol. 10, no. 1, p. 2539, Feb. 2020. 5, 8
- [46] S. Wei, D. Bloemers, and A. Rovira, "A Preliminary Study of the Eye Tracker in the Meta Quest Pro," in *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*, ser. IMX '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 216–221. 5
- [47] M. Trinidad-Fernández, B. Bossavit, J. Salgado-Fernández, S. Abbate-Chica, A. J. Fernández-Leiva, and A. I. Cuesta-Vargas, "Head-Mounted Display for Clinical Evaluation of Neck Movement Validation with Meta Quest 2," *Sensors*, vol. 23, no. 6, p. 3077, Jan. 2023. 5
- [48] D. O. Hebb, *The Organization of Behavior; a Neuropsychological Theory*, ser. The Organization of Behavior; a Neuropsychological Theory. Oxford, England: Wiley, 1949. 5
- [49] L. Guadron, A. J. van Opstal, and J. Goossens, "Speed-accuracy trade-offs influence the main sequence of saccadic eye movements," *Scientific Reports*, vol. 12, no. 1, p. 5262, Mar. 2022. 6
- [50] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*. Oxford University Press, Jun. 2015. 6
- [51] C. Doersch, Y. Yang, M. Vecerik, D. Gokay, A. Gupta, Y. Aytar, J. Carreira, and A. Zisserman, "TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement," Jun. 2023. [Online]. Available: <https://arxiv.org/abs/2306.08637v2> 8
- [52] X. Chen, L. Lu, and H. Wei, "Identifying Fixations and Saccades in Virtual Reality," in *Proceedings of the 2024 International Conference on Virtual Reality Technology*, ser. ICVRT '24. New York, NY, USA: Association for Computing Machinery, Mar. 2025, pp. 24–31. 8
- [53] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Optical Science and Technology, SPIE's 48th Annual Meeting*, B. Bosacchi, D. B. Fogel, and J. C. Bezdek, Eds., San Diego, California, USA, Jan. 2004, p. 64. 8
- [54] H. Finch, "An introduction to the analysis of ranked response data," vol. 27, no. 1. [Online]. Available: <https://openpublishing.library.umass.edu/pare/article/id/1331/> 8
- [55] S. Andrist, B. Mutlu, and M. Gleicher, "Conversational Gaze Aversion for Virtual Agents," in *Intelligent Virtual Agents*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, vol. 8108, pp. 249–262. 9
- [56] R. McDonnell, M. Breidt, and H. H. Bülthoff, "Render me real?: Investigating the effect of render style on the perception of animated virtual humans," vol. 31, no. 4, pp. 1–11. [Online]. Available: <https://dl.acm.org/doi/10.1145/2185520.2185587> 9